FUSION PROTEIN AND USES THEREOF

FIELD OF THE INVENTION

The present invention is directed to fusion proteins containing a single chain antibody (sFv) in place of a modular protein binding domain, or an antibody epitope in place of the linear binding epitope for a modular protein binding domain and the use in an assay system for (1) validating that a protein-protein interaction causes a specific biological activity, (2) identifying target molecules capable of affecting those interactions and (3) identifying the biological activities involved in such interactions.

BACKGROUND OF THE INVENTION

A wide range of cellular processes involve protein-protein interactions. For example, the interaction between a receptor and its target ligand (such as $VEGF_{165}$ Receptor - Neuropilin-1 complexes) as well as intracellular interactions (such as adaptor-kinase complexes). These interactions can cause signal transduction or processes governing whether a cell will proliferate, differentiate, die, adhere, migrate or otherwise respond to its environment.

These interactions typically involve the modular protein binding domains (MPBD) of one protein, which are regions of about 60 to 200 amino acids, and the corresponding binding site of the second protein. Examples of such domains are SH2 (src homology 2), SH3 (src Homology 3) and PTB. These domains typically bind to linear peptide epitopes of about 4-10 amino acid residues on their binding partners.

These MPBDs are present in a wide variety of functionally distinct proteins. The SH2 domain, which binds to phosphorylated tyrosine residues, appears to be associated with a wide range of activated growth factors, e.g. epidermal growth factor, platelet-derived growth factor, fibroblast growth factor, etc.

Despite the fact that numerous MPBDs and their corresponding binding sites (BSs) have been identified, however, this typically does not lead to

identification of whether a pair of proteins actually interact *in vivo* and whether that interaction is important for biological activity. This is because the binding interaction between MPBDs and their corresponding binding sites is not highly specific, but is only partially specific. Consequently, one protein can bind with several to hundreds of other proteins with virtually the same affinity. This lack of specificity has created a fundamental stumbling block. For example, if a single SH3 domain can bind with virtually identical affinity to tens or even hundreds of different proteins in the cell, and a single SH3 binding site can combine with tens or hundreds of different SH3 domains, which of these hundreds of potential protein complexes actually results in a specific *in vivo* function? For example, if protein A binds with about equal affinity to proteins 1-200, it would be exceedingly difficult to determine whether the complex of protein A-protein 5 and not the complex of protein A-protein55 causes a specific behavior. Eliminating all 200 proteins from one's test system, will likely introduce numerous undesired cellular functional artifacts.

Thus, trying to determine the specific pair of protein-protein interactions that results in a biological activity is typically done indirectly. For instance, one can try to accomplish this identification negatively, i.e. by preventing the specific interaction between the proteins to determine if preventing binding of a set of proteins eliminates a function. This would be done by altering the MPBD and/or BS to prevent binding. However, such an alteration will also prevent binding with other proteins. Therefore that method does not definitively confirm that the specific biomolecular protein complex under investigation actually interacts in the cell to result in the function eliminated. For example, protein A may interact with protein 55 at one point in a pathway, whereas protein 5 interacts with protein D at a different point in the pathway. Thus, by altering the binding sites of proteins A and 5, the function could be lost without the two proteins having to interact directly with each other.

It would be desirable to have a protein that can still perform its native functions, but can bind more specifically to putative partners. It would also be useful to have a method to replace the relatively nonspecific interactions

between two proteins with a highly specific interaction, thereby allowing the two proteins of interest to directly interact without concern for competing interactions with other proteins in the cell.

Another indirect approach used are "interaction trap" systems. For example, the yeast two hybrid approach. In this method a selectable output such as growth on a selected media, or metabolism of colorimetric substrates is dependent on reconstituting a protein-protein interaction with a "bait" protein. Such a system is limited, in that the selectable biological output is fixed by the experimental system and proteins are screened for their ability to bind to the target (bait) protein. Further, one is typically trying to reconstitute function in a foreign system, e.g. yeast, as opposed to a mammalian cell system. It would be desirable to have a system that forces interactions between the proteins and looks at their functional consequences. It would also be desirable to have a system that more closely resembles the actual cellular microenvironment where the protein-protein interactions occurs.

SUMMARY OF THE INVENTION

We have discovered novel fusion proteins that can be used in assays to identify protein-protein interactions. One fusion protein comprises a protein containing a modular protein binding domain (MPBD), wherein the MPBD is substituted by a single chain antibody. Preferably, the MPBD is selected from the group of domains consisting of src homology 2 (SH2), src homology 3 (SH3) phosphotyrosine binding (PTB), WW, PDZ, 14.3.3, WD40, EH, Lim, etc.

For example, such a protein is a tyrosine kinase.

A second fusion protein comprises a protein containing a binding site that binds to a modular protein binding domain MPBD, wherein at least one linear epitope that binds to the MPBD within the binding site is substituted by an antigenic epitope of 6-20 amino acids that binds to the single chain antibody that has been substituted for the MPBD. Preferably, the second fusion protein

contains multiple copies of the epitope. For example, 2-20, more preferably 3-15, still more preferably 4-10 copies of the epitope.

Nucleic acid sequences encoding these fusion proteins can be prepared by known techniques. Preferably these sequences (e.g., genes) are contained in vectors and are operably linked to a promoter.

These vectors can be used to transform a cell.

These transformed cells can be used to identify the function of a protein-protein interaction, to identify a particular protein involved in an interaction and to study the specific effect of specific functional domains. In one embodiment there is an assay for determining the activity of a protein-protein interaction, comprising:

(a) transforming a cell by a vector containing a gene encoding the protein wherein the MPBD site has been substituted by single chain antibody and a second vector, wherein said second vector contains a gene encoding a fusion protein that has a binding site for a MPBD, wherein a linear epitope of 6-20 amino acids that is bound specifically by the single chain antibody is substituted for the binding site that binds to MPBD;

(b) culturing the transformed cell;

(c) and comparing the activity to a base line control; and

(d) measuring changes in activity to determine the activity caused by that protein-protein interaction.

Preferably, the cell used does not express the protein containing the MPBD of the fusion protein encoded by the gene contained in the transforming vector or the effect of the interaction is dominant or assayed in a way that does not depend on the lack of the wild-type counterpart of the engineered gene. The control can be an untransformed cell, or more preferably cells transformed with each of the modified fusion proteins, alone, but not together. By this latter way, one can determine the effect of expression of each of the proteins on the cell and determine what effects are dependent on the interaction of the two proteins.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the nucleotide sequence (SEQ ID NO:1) and the amino acid sequence (SEQ ID NO:2) of 3DX, a second-generation ScFv derived from monoclonal antibody 9E10.

Figures 2A and 2B show the antibody binding of antibody 3DX to a Myc-tagged protein (Myc_6 NWASP) for proteins bound to beads ($GSH_{pull\ down}$) (Figure 2A) or from whole cell lysates (Figure 2B).

DETAILED DESCRIPTION OF THE INVENTION

We have now discovered that by creating a first fusion protein wherein a single-chain antibody is substituted for a modular protein binding domain (MPBD) of sixty to two hundred amino acids and a second fusion protein wherein the linear epitope of the corresponding binding site for that modular protein binding domain has been substituted with at least one copy of an epitope to the single-chain antibody, one can obtain high affinity protein-protein binding interactions, to the exclusion of competing reactions. Moreover, these changes can be tailored so that one does not change or minimally changes the conformation of the fusion protein in comparison to the wild type conformation.

Consequently, one does not have to be concerned about competition resulting in different interactions between one of the proteins and an unknown protein or that the change results in a functionally inactive protein.

Accordingly, one can now assay for the specific affect that occurs when protein A binds with protein B. This contrasts with previous methodologies where in trying to determine the effect of such an interaction, one was looking at negative effects, i.e. deleting the MPBD of protein A and the binding site protein of B. Such an assay could not clearly establish that protein A actually interacts with protein B. For example, assume protein A interacts with protein D and protein B interacts with protein F and that the interaction is in the same biological pathway. Deletions in the MPBD in protein A would prevent binding with protein E, which could result in a loss of activity. Similarly a deletion in

the binding site of protein B would prevent its ability to bind with protein E, thereby also resulting in a loss of activity. However, by looking at the change in function occurring from the above-mentioned deletions, one would not know that protein A and protein B did not actually interact with each other.

By contrast, with the fusion proteins created by inserting a single chain antibody in place of the MPBD of the protein A, the ability of protein A to interact with protein D would be eliminated. And by inserting at least one antigen epitope in place of the linear binding site epitope of protein B, the ability of protein B to interact with protein E is eliminated. Instead only protein A and protein B can interact with each other. Moreover, since they act positively, namely by actually interacting one can look for generation of a function instead of loss of function for the complex. And by retaining the overall conformation of fusion protein A and wild type protein A and fusion protein B and wild type protein B, their activity remains the same as the wild type proteins. Thus, one can determine what activity actually occurs *in vivo* when protein A and protein B interact. We sometimes refer to this screening methodology as the "Functional Interaction Trap" (FIT).

One can transform any cell with vectors containing genes encoding the fusion proteins. One can then compare these fusion protein transformed cells with a base line of a control of the same cell to see the differences that occur in these cells. Preferably, the genes would be under the control of a promoter that results in high levels of expression of the fusion protein. Other instances and sequences that result in enhancing expression can be used. For example, a HIV TAR element upstream and operably linked to the gene and a sequence encoding the HIV tat will result in increasing expression by factors of 10^3. Other such sequences are known.

Preferably, one can use a cell line where at least one of the native proteins being studied is not expressed or minimally expressed. Such cells may occur naturally or may be engineered. These cells are referred to herein as "knockout" cells.

Knockout cells are well known in the art. Many knock out cells are commercially available from a variety of different manufacturers. If one desires, one can reduce the level of endogenous gene expression of a particular protein in a cell i.e. knock out the gene by using anti-sense, or ribozyme approaches to inhibit or prevent translation of the protein's mRNA transcripts, preferably targeting the endogenous promoter; intracellular targeting of antibodies to the protein to prevent its expression, triple helix approaches to inhibit transcription of the gene, targeted homologous recombination to inactivate or knock out the gene or the endogenous promoter. Preferably one would knock out the endogenous promoter. See Wagner, R. *Nature*, 3782:333-335 (1994), Saver et al., *Science*, 247:1222-1225 (1990), Helene, C., *Anti-Cancer Drug Des.*, 6:569-84 (1991); Helene, C. et al., *Ann. NY Acad. Sci.*, 660:27-36 (1992); Marasco, W. (US Patent No. 5,965,374). Alternatively, as mentioned above, there are many instances where the protein of interest is not being expressed at a particular time in a cell's cycle or in a particular type of cell. Such cells can readily be screened for, for example, by looking for presence of the transcript of the protein using standard detection means including differential display techniques.

Additionally, in many cases the phenotype will be dominant (such as seen with oncogenes, viral proteins, etc), and the activity caused by the protein-protein interaction will be seen in cells which still express the two proteins.

In creating one of the fusion proteins, the MPBD of the protein containing it is replaced with a single chain antibody (sFv). sFv's are typically about 250 amino acids in length. Thus, they can readily be substituted for the MPBD without significantly affecting the conformation of the protein. One can readily tailor the specific sFv used based on the general conformation of the protein, and the specific conformation of the MPBD. Single chain antibodies are well known in the art (see U.S. Patent No. 4,946,778) and also commercially available. Preferred single chain antibodies that can be used in the present invention include those based on 12CA5, which recognizes the influenza virus hemagglutinin (HA) epitope, and 9E10, which recognizes an epitope from the

Myc protein. Care should be taken that the epitope that the antibody binds to does not result in a reaction in the cell that can affect the behaviors being monitored. This can readily be determined by known techniques. For example, some epitopes are known to be far more antigenic than others. Alternatively, one can test a cell by transforming it with a vector containing a gene encoding the fusion protein expressing the epitope of interest and determining if this transformation adversely affects the cell as compared to a control baseline. By using standard techniques, one can select antibodies that have the desired characteristics. For example, using phage display methodologies [Burton, D.R., et al., *Proc Natl Acad Sci* USA 88:10134-10137 (1991); Hoogenboom H.R., et al., *Immunol Rev* 130:41-68 (1992); Winter G., et al., *Annu Rev Immunol* 12:433-455 (1992); Marks, J.D., et al., *J Biol Chem* 267:16007-16010 (1992); Nissim, A., et al., *EMBO J* 13:692-698 (1994); Vaughan T.J., et al., *Nature Bio* 14:304-314 (1996); Marks C., et al., *New Eng J Med* 335:730-733 (1996)]. For example, very large naïve human sFv libraries have been and can be created to offer a large source or rearranged antibody genes against a plethora of target molecules. Smaller libraries can be constructed from individuals with a particular characteristic such as autoimmune [Portolano S., et al., *J Immunol* 151:2839-2851 (1993); Barbas S.M., et al., *Proc Natl Acad Sci USA* 92:2529-2533 (1995)] or infectious diseases [Barbas C.F., et al., *Proc Natl Acad Sci USA* 89:9339-9343 (1992); Zebedee S.L., et al., *Proc Natl Acad Sci USA* 89:3175-3179 (1992)] in order to isolate disease specific antibodies. Another such construction includes the use of human monoclonal antibody producing cell lines. [Marasco, W.A., et al., *Pro Natl Acad Sci USA*, 90:7889-7893 (1993); Chen, S.Y., et al., *Proc Natl Acad Sci USA* 91:5932-5936 (1994). *In vitro* procedures to manipulate the affinity and fine specificity of the antigen binding site have been reported including repertoire cloning [Clackson, T., et al., *Nature* 352:624-628 (1991); Marks, J.D., et al., *J Mol Biol* 222:581-597 (1991); Griffiths, A.D., et al., *EMBO J* 12:725-734 (1993)], *in vitro* affinity maturation [Marks, J.D., et al., *Biotech* 10:779-783 (1992); Gram H., et al., *Proc Natl Acad Sci USA* 89:3576-3580 (1992)], semi-synthetic libraries [Hoogenboom, H.R., *supra*; Barbas, C.F., *supra*; Akamatsu, Y., et al., *J Immunol* 151:4631-4659 (1993)] and guided selection [Jespers, L.S., et al., *Bio Tech* 12:899-903 (1994)]. One can subject an sFv to *in*

vitro mutagenesis and select for variants having the desired characteristics such as improved binding.

For example, one can take the 9E10 and 12CA5 antibody, subject it to *in vitro* mutagenesis, screen a library such as a phage display library. One can test for binding by using an assay such as an ELISA-based assay system. One can create fusions of the mutant sFvs to, for example, glutathione.S-transferase (GST) and expressing the GST fusions in a transiently transfected cell line such as 293T cells along with various proteins tagged with the desired epitope, e.g., Myc or HA epitopes. The antibodies are compared against the parent antibody. One antibody, derived from 9E10, displayed dramatically improved binding, termed 3DX (see SEQ ID NOS 1 and 2 and Figure 1).

The novel fusion proteins of the present invention can be used in a number of ways. For example, in an assay to determine the biological activity of a specific protein-protein pair.

The following example is illustrative of the ways one can look at a protein-protein interaction. The role of adaptor proteins is to modulate the localization, local concentration and binding partners of the proteins with which they interact. The fundamental problem with fully understanding their activity is that they contain MPBD, -- namely, SH2 and SH3 domains-- that are promiscuous in their binding activities (i.e. are not specific). Abl is a kinase that has a corresponding binding site that is believed to form a complex with an adaptor protein. Yet, because of the mutual lack of specificity for specific adaptor-kinase complexes it is difficult to assess which of the many adaptors that Abl can bind to *in vivo* is responsible for the activity of interest. This is because Abl can bind to the SH3 domains of a range of adaptors. Similarly, the multiple adaptors that contain the SH3 domains that can bind to the Abl-adaptor-binding site may also bind to numerous proteins other than Abl.

By the present invention, the interaction interface between the SH3 domain of the adaptor and their binding sites on Abl (the kinase) can be replaced

by a synthetic high affinity, specific interaction i.e. sFv and its ~10 amino acid epitope. Preferably the main SH3 domain in an adaptor is replaced with sFv and at least one of the corresponding sFv epitope is inserted into an Abl molecule in which the authentic adaptor binding sites have been ablated. This results in only one interaction: sFv in place of the SH3 domain of adaptors (e.g. Crk, Nck, Grb2) + sFv epitope-Abl. Preferably one inserts multiple copies of the epitope to increase the avidity of interaction between the sFv and the tagged partner. For example, 2-20 copies. Preferably, 3-15 copies of the epitope. Still more preferably 4-10 copies of the epitope. Even more preferably, 4-6 copies.

The addition of spacers and/or linkers between the epitope or the sFv and their fusion partner can help maintain a wild-type conformation. Preferred linkers include neutral amino acid residues such as Gly. The biological activity of the synthetic/kinase adaptor pairs are assayed in standard tissue culture models. Accordingly, the application of the invention allows a recombinant adaptor and kinase to specifically bind to each other in the absence of competing interactions.

In another embodiment, the fusion protein, e.g. synthetic adaptor proteins, can be based on a sFv that is used to bring specified functional groups into close proximity to the Abl catalytic domain. Various functional domains for a wide range of proteins are well known in the art. Other functional domains can readily be determined by known means such as the deletion mutant technique. Virtually any arrangement of functional domains desired can be prepared. These proteins can be used for a range of functions, for example to:

1. test significance of interaction between a specific pair of proteins (one with MPBD, one with BS) *in vitro*, in cultured cells, or in genetically modified animals.

2. test which, of a series of potential protein interaction partners for a protein of interest identified by a prior screening assay, such as the yeast two-

hybrid assay, is responsible for biological activity of interest *in vitro*, in cultured cells, or in genetically modified animals;

3. to identify binding partners for protein of interest containing MPBD or BS whose binding, directed by sFv/epitope interactions results in biological activity of interest in cells (e.g., screen libraries containing sFv or epitope in cells expressing the protein of interest modified with epitope or sFv).

4. to identify proteins that, when associated with another protein-based functional group by sFv/epitope-mediated interactions, leads to biological activity of interest (Functional groups include subcellular targeting signals, oligomerization domains, engineered constructs permitting induced aggregation, etc. These can include functional groups fused to sFv or epitope, screened with libraries of proteins containing epitope or sFv); and

5. to append functional groups (such as chemical groups, fluorescent dyes, etc.) to protein or proteins of interest in cells or *in vitro* (Such functional groups can be covalently coupled to synthetic peptide epitope and proteins of interest modified with sFv by known means).

More specifically, one can create ectopic SH2 domains. For example one method of bringing novel functional domains into proximity with a target protein is by using the N-terminal SH2 of GAP, which is known to have a high affinity for phosphorylated p62dok. Thus, one can determine if Abl facilitates phosphorylation (measurable biological output) of p62dok (second known protein). Adaptors bind Abl via their SH3 domains, and provide an SH2 domain to the kinase, which may be useful for phosphorylation of some proteins by Abl. Using this system novel SH2 domains such as that of GAP can be brought into close proximity to Abl. In this way, one can, for example, look at the effect of the specific phosphorylation of a single protein by a kinase. While kinases can have profound biological effects, it is not clear what the consequences of phosphorylation of any single substrate might be. For example, Abl can transform cells, but the specific substrates essential for this product are not

known. However, with Abl oncogenic transformation requires a functional SH2 domain [Mayer, B.J., et al., *Mol. Cell. Biol.* 12:609-610 (1992). Accordingly, if the SH2 domain of Abl is replaced with a sFv, the mutant will not transform cells or phosphorylate substrates efficiently. Then, using at least one copy of an epitope to the sFv in a putative protein or a library of proteins, one can identify the protein that Abl specifically phosphorylated. In this manner, one can evaluate the biological consequences of phosphorylation of the specific protein, or to isolate and identify from a library, those substrates (proteins) displaying a desired property (e.g., transformation). Other proteins involved in phosphorylation in addition to Abl can readily be used.

Similarly, one can create specific targeting domains such as the focal adhesion target (FAT) region of FAK. This allows the testing of effects of relocalization (or targeting) in the absence of the increased processitivity conferred by an SH2 domain.

In order to circumvent the lack of specificity of SH2-phosphopeptide interactions, e.g. an SH2/SH3 adaptor, for instance the Nck SH3 domains can be directly targeted to subcellular locations known to directly harbor Nck SH2-binding sites, allowing for the experimental mimicking of signal-induced creation of localized binding sites for the Nck SH2 domain without relocalizing other SH2-containing proteins.

By appropriate mutagenesis one can alter the local concentration of Nck SH3 domains on the membrane, mimic clustering of the tyrosine phosphorylated sites involved in normal stimuli such as TCR engagement or receptor clustering e.g. Eph receptor.

Accordingly one can replace any MPBD, such as the SH3 domain or domains of for example Nck with the V_L and V_H of a known antibody. These sFv's can be constructed by standard techniques for example PCR such as RT-PCR from mRNA isolated from a monoclonal antibody's hybridoma cells using protocols and reagents in commercially available kits such as those available

from Pharmacia Biotech supplemented with primers whose design can be based on published reports. The V_H and V_L, i.e. the sFv fragments can be designed with 5' and 3' restriction sites allowing easy fusion of the sFv downstream of a specified site such as glutathione S-transferase (GST) in a gene, which can be in a desired expression vector, for example, the mammalian expression vector pEBG [Mayer, et al., *Current Biol.*, 5:296-305 (1995)]

Consequently, one can use the fusion sFv/epitope combinations described here as a screening technology to validate suspected interactions and to identify functionally important interations.

PCR generated clones can be tested for their ability to co-precipitate the epitope protein, for example, in a transfected mammalian cell. This permits the identification of the appropriate single chain antibody clones that have the desirable binding. The sFv identified can be transferred by PCR based mutagenesis into the gene encoding the desired clone e.g. Nck, by replacing the MPBD such as SH3 most important for the protein binding e.g. SH3-2. The remaining SH3 encoding domains in that gene can be mutated to prevent interaction with endogenous proteins.

The corresponding binding motifs in the protein of interest can also be generated by mutagenesis to have the corresponding epitope to the single chain antibody in place of the binding portion of the binding site. For example, Nck binds to Abl. This binding is mediated by 3 PxxP SH3-binding motifs in Abl located C-terminal to the catalytic domain. The most important of these motifs is at position 630 in murine c-Abl. In order to minimally affect that protein the entire binding region should not be altered. Instead one mutates the minimal amount necessary. Preferably, changing the two minor binding sites by site directed mutagenesis such as substituting Ala residues for Pro. Thereafter, the antibody's epitope is inserted. For example, the epitope to the 12CA5 antibody is (SEQ ID NO:3) YPYDVPDYA and this can be inserted within the third PxxP motif to create the desired epitope. The appropriate changes can readily be made depending on the particular protein being used. For example Cbl binding

to Nck is mediated by both a large Pro rich region between amino acids 482 and 689 of human c-Cbl and a short motif PERPPKP (SEQ ID NO:4) between 819 and 827. Thus, one would insert, for example, the Myc or the HA epitope in place of the second binding site and alter or delete the larger Pro rich region. These altered proteins can be constructed by a variety of methods. For example, two step PCR based mutagenesis using Pfu polymerase to minimize polymerase errors.

Constructs can be sequenced in the area of primer binding site to confirm the presence of mutations and to detect any unforeseen changes. This methodology can be used with virtually any protein pair.

Even if one can not readily test protein pairs in a cellular system to ensure that it is the binding between these two proteins that is responsible for a particular function, one can look at such prospective native proteins in any system and tag them and use deletion mutants to determine which are the most important MPBDs and binding sites in those two proteins.

This approach can be adapted for studying a wide range of protein-protein interactions. For example, the interaction of cellular proteins with certain viral proteins can have many undesired affects. For instance it is known that the HIV Nef protein interacts with cellular protein(s) to reduce immune function. However the interaction of Nef with particular protein(s), where binding has functional consequences, are not currently known because of the lack of specificity in protein binding. One could take a Nef protein insert a particular epitope into its binding site and attach a single chain antibody, specific for the inserted epitopes, to a library of proteins. Thereafter, one can transform cells with the Nef protein and the library of different proteins. In this manner one can rapidly identify the protein that interacts with Nef to result in the function being sought. Once the particular protein is identified, which can be done by standard techniques, e.g., by downregulation of immune function by Nef, one can look for compounds that interact with this protein. This permits a rational drug design.

Alternatively, one can take a protein that has a MPBD and replace that with a specific single chain antibody and have the unknown library of proteins that preferably contain an epitope that binds to the MPBD tagged to (fused to) the epitope for the sFv. Preferably, one would replace the MPBD binding site with the antibody, but typically the proteins will be tagged randomly. In an alternative embodiment, the library can be a library of proteins containing the MPBDs, wherein the proteins are tagged to a single chain antibody and this library is used to determine the effect of a protein binding to a protein containing an MPBD binding site that has been replaced with an antibody epitope. Typically, one will use a library of nucleic acid sequences encoding the desired proteins. These nucleic acid libraries will be used to create a vector library that can be used to transform the desired cells.

Alternatively, fusing the sFv or the epitopes for the sFv to various dominant targeting sequences or other functional domains e.g. membrane localization, nuclear localization, focal adhesions or cross-linkable membrane domains allows for screening for libraries of either epitope target fusion proteins or sFv target antibodies fusion proteins. One can then screen for a specific biological output in specific cellular compartments. Thereafter, one can use standard techniques to identify the protein complex.

One can also look at cells surface receptors to better understand cell to cell interaction.

One can also look at a particular interaction between two proteins isolated from a cell, for example, to look for molecules that interact with these proteins.

As aforesaid, in one embodiment, one can bring various functional groups in association with a protein of choice. This requires that one of the two proteins be fused to a larger protein domain, i.e. the sFv, with the second protein only fused to a small 6-20 amino acid residue epitope, more preferably 6-10 amino

acid residue epitope, and still more preferably 10 amino acid residue epitope for the single chain antibody. This results in the association of specific domains without the problem of having to use large proteins which can place structural and sterical constraints on the fusion partner.

Many antibodies and sFv's are known in the art and readily available. For example, 1287 human V_H sequences and 1041 human V_L sequences are available in Andrew C.R. Martin's Kabat Man web page (http://www.biochem.ucl.ac.uk/~martin/abs/simkab.html).

Alternatively one can readily prepare a desired single chain antibody. One method is by using hybridoma mRNA or splenic mRNA as a template for PCR amplification of such genes [Huse, et al., *Science* 246:1276 (1989)]. For example, antibodies can be derived from murine monoclonal hybridomas [Richardson J.H., et al., *Proc Natl Acad Sci USA* 92:3137-3141 (1995); Biocca S., et al., *Biochem and Biophys Res Comm*, 197:422-427 (1993) Mhashilkar, A.M., et al., *EMBO J.* 14:1542-1551 (1995)]. These hybridomas provide a reliable source of well-characterized reagents for the construction of antibodies and are particularly useful when their epitope reactivity and affinity has been previously characterized. Another source for such construction includes the use of human monoclonal antibody producing cell lines. [Marasco, W.A., et al., *Proc Natl Acad Sci USA*, 90:7889-7893 (1993); Chen, S.Y., et al., *Proc Natl Acad Sci USA* 91:5932-5936 (1994)]. As discussed above, another method includes the use of antibody phage display technology to construct new antibodies against different epitopes on a target molecule.

Other sources include transgenic mice that contain a human immunoglobulin locus instead of the corresponding mouse locus as well as stable hybridomas that secrete human antigen-specific antibodies. [Lonberg, N., et al., *Nature* 368:856-859 (1994); Green, L.L., et al., *Nat Genet* 7:13-21 (1994)]. Such transgenic animals provide another source of human antibody genes through either conventional hybridoma technology or in combination with phage display technology. Starting materials for these recombinant DNA based strategies

include RNA from mouse spleens [Clackson, T., *supra*] and human peripheral blood lymphocytes [Portolano, S., et al., *supra*; Barbas, C.F., et al., *supra*; Marks, J.D., et al., *supra*; Barbas, C.F., et al., *Proc Natl Acad Sci USA* 88: 7978-7982 (1991)] and lymphoid organs and bone marrow from HIV-1-infected donors [Burton, D.R., et al., *supra*; Barbas, C.F., et al., *Proc Natl Acad Sci USA* 89:9339-9343 (1992)].

Accordingly, antibody genes can be prepared based upon the present disclosure by using any known techniques such as those described.

Thereafter, using any of these antibodies, one can construct V_H and V_L genes. For instance, one can create V_H and V_L libraries from murine spleen cells that have been immunized either by the above-described *in vitro* immunization technique or by conventional *in vivo* immunization and from hybridoma cell lines that have already been produced or are commercially available. One can also use commercially available V_H and V_L libraries. One method involves using the spleen cells to obtain mRNA which is used to synthesize cDNA. Double stranded cDNA can be made by using PCR to amplify the variable region with a degenerative N terminal V region primer and a J region primer or with V_H family specific primers, e.g., mouse-12, human-7.

For example, the genes of the V_H and V_L domains of the desired antibody such as one to influenza hemaglutin can be cloned and sequenced. The first strand cDNA can be synthesized from, for example, total RNA by using oligo dT priming and the Moloney murine leukemia virus reverse transcriptase according to known procedures. This first strand cDNA is then used to perform PCR reactions. One would use typical PCR conditions, for example, 25 to 30 cycles using e.g. Vent polymerase or Pfu polymerase to amplify the cDNA of the immunoglobulin genes. DNA sequence analysis is then performed. [Sanger, et al., *Proc. Natl. Acad. Sci. USA* 79:5463-5467 (1977)].

Both heavy chain primer pairs and light chain primer pairs can be produced by this methodology. One preferably inserts convenient restriction

sites into the primers to make cloning easier. The V_H and V_L chains can be joined together by convenient linkers, which are known in the art.

Those of ordinary skill in the art will recognize that a large variety of possible moieties can be coupled to the resultant antibodies or to other molecules of the invention for ease in subsequently identifying a functionally important complex. This is particularly useful when one is using a library of proteins to identify the unknown protein that interacts with a known protein. See, for example, "Conjugate Vaccines", Contributions to Microbiology and Immunology, J.M. Cruse and R.E. Lewis, Jr (eds), Carger Press, New York, (1989), the entire contents of which are incorporated herein by reference. In some instances, one would use genetic engineering to couple moieties.

Coupling may be accomplished by any chemical reaction that will bind the two molecules so long as the antibody and the other moiety retain their respective activities. This linkage can include many chemical mechanisms, for instance covalent binding, affinity binding, intercalation, coordinate binding and complexation. The preferred binding is, however, covalent binding. Covalent binding can be achieved either by direct condensation of existing side chains or by the incorporation of external bridging molecules. Many bivalent or polyvalent linking agents are useful in coupling protein molecules, such as the antibodies of the present invention, to other molecules. For example, representative coupling agents can include organic compounds such as thioesters, carbodimides, succinimide esters, diisocyanates, glutaraldehydes, diazobenzenes and hexamethylene diamines. This listing is not intended to be exhaustive of the various classes of coupling agents known in the art but, rather, is exemplary of the more common coupling agents. (See Killen and Lindstrom 1984, "Specific killing of lymphocytes that cause experimental Autoimmune Myasthenia Gravis by toxin-acetylcholine receptor conjugates." *Jour. Immun.* 133:1335-2549; Jansen, F.K., H.E. Blythman, D. Carriere, P. Casella, O. Gros, P. Gros, J.C. Laurent, F. Paolucci, B. Pau, P. Poncelet, G. Richer, H. Vidal, and G.A. Voisin. 1982. "Immunotoxins: Hybrid molecules combining high specificity and potent cytotoxicity". *Immunological Reviews* 62:185-216; and Vitetta et al., supra).

Preferred linkers are described in the literature. See, for example, Ramakrishnan, S. et al., *Cancer Res.* 44:201-208 (1984) describing use of MBS (M-maleimidobenzoyl-N-hydroxysuccinimide ester). See also, Umemoto et al. U.S. Patent 5,030,719, describing use of halogenated acetyl hydrazide derivative coupled to an antibody by way of an oligopeptide linker. Particularly preferred linkers include: (i) EDC (1-ethyl-3-(3-dimethylamino-propyl) carbodiimide hydrochloride; (ii) SMPT (4-succinimidyloxycarbonyl-alpha-methyl-alpha-(2-pyridyl-dithio)-toluene (Pierce Chem. Co., Cat. (21558G); (iii) SPDP (succinimidyl-6 [3-(2-pyridyldithio) propionamido] hexanoate (Pierce Chem. Co., Cat #21651G); (iv) Sulfo-LC-SPDP (sulfosuccinimidyl 6 [3-(2-pyridyldithio)-propianamide] hexanoate (Pierce Chem. Co. Cat. #2165-G); and (v) sulfo-NHS (N-hydroxysulfo-succinimide: Pierce Chem. Co., Cat. #24510) conjugated to EDC.

The linkers described above contain components that have different attributes, thus leading to conjugates with differing physio-chemical properties. For example, sulfo-NHS esters of alkyl carboxylates are more stable than sulfo-NHS esters of aromatic carboxylates. NHS-ester containing linkers are less soluble than sulfo-NHS esters. Further, the linker SMPT contains a sterically hindered disulfide bond, and can form conjugates with increased stability. Disulfide linkages, are in general, less stable than other linkages because the disulfide linkage is cleaved in vitro, resulting in less conjugate available. Sulfo-NHS, in particular, can enhance the stability of carbodimide couplings. Carbodimide couplings (such as EDC) when used in conjunction with sulfo-NHS, forms esters that are more resistant to hydrolysis than the carbodimide coupling reaction alone.

Antibodies that are part of the fusion proteins of the present invention can be detected by appropriate assays, e.g., conventional types of immunoassays. For example, a sandwich assay can be performed in which an antibody to the antibody or a specific fragment thereof is affixed to a solid phase. Incubation is maintained for a sufficient period of time to allow the antibody in the sample to bind to the immobilized antibody on the solid phase. After this first incubation,

the solid phase is separated from the sample. The solid phase is washed to remove unbound materials and interfering substances such as non-specific proteins which may also be present in the sample. The solid phase containing the antibody of interest bound to the immobilized polypeptide is subsequently incubated with labeled antibody or antibody bound to a coupling agent such as biotin or avidin. Labels for antibodies are well-known in the art and include radionucleotides, enzymes (e.g. maleate dehydrogenase, horseradish peroxidase, glucose oxidase, catalase), fluors (fluorescein isothiocyanate, rhodamine, phycocyanin, fluorescamine), biotin, and the like. The labeled antibodies are incubated with the solid phase and the label bound to the solid phase is measured, the amount of the label detected serving as a measure of the amount of the antibody of interest present in the sample. These and other immunoassays can be easily performed by those of ordinary skill in the art.

The resultant fusion proteins can be expressed by a vector containing a DNA segment encoding the single chain antibody-MPBD containing protein described above.

These can include vectors, liposomes, naked DNA, adjuvant-assisted DNA, gene gun, catheters, etc. Vectors include chemical conjugates such as described in WO 93/04701, which has targeting moiety (e.g. a ligand to a cellular surface receptor), and a nucleic acid binding moiety (e.g. polylysine), viral vector (e.g. a DNA or RNA viral vector), fusion proteins such as described in PCT/US 95/02140 (WO 95/22618) which is a fusion protein containing a target moiety (e.g. an antibody specific for a target cell) and a nucleic acid binding moiety (e.g. a protamine), plasmids, phage, etc. The vectors can be chromosomal, non-chromosomal or synthetic.

Preferred vectors include viral vectors, fusion proteins and chemical conjugates. Retroviral vectors include moloney murine leukemia viruses. Another vector is a lentiviral vector. Preferably the lentiviral vector is a pseudotyped lentiviral vector, which contains heterologous envelope glycoprotein. See, e.g., U.S. Patent No. 5,981,276. One can use these fusion

proteins to examine the effect of their interactions in *in vivo* systems. Thus, vectors can be selected depending upon the cells where the interaction is being examined. These vectors include pox vectors such as orthopox or avipox vectors, herpesvirus vectors such as a herpes simplex I virus (HSV) vector [Geller, A.I. *et al.*, *J. Neurochem*, 64: 487 (1995); Lim, F., *et al.*, in DNA Cloning: *Mammalian Systems, D. Glover, Ed. (Oxford Univ. Press, Oxford England) (1995)*; Geller, A.I. *et al.*, *Proc Natl. Acad. Sci.*: U.S.A. 90:7603 (1993); Geller, A.I., *et al.*, *Proc Natl. Acad. Sci USA* 87:1149 (1990)], Adenovirus Vectors [LeGal LaSalle et al., *Science*, 259:988 (1993); Davidson, *et al.*, *Nat. Genet* 3:219 (1993); Yang, *et al.*, *J. Virol.* 69:2004 (1995)] and Adeno-associated Virus Vectors [Kaplitt, M.G., *et al.*, *Nat. Genet.* 8:148 (1994)].

Pox viral vectors introduce the gene into the cells cytoplasm. Avipox virus vectors result in only a short term expression of the nucleic acid. Adenovirus vectors, adeno-associated virus vectors and herpes simplex virus (HSV) vectors are preferred for introducing the nucleic acid into neural cells. The adenovirus vector results in a shorter term expression (about 2 months) than adeno-associated virus (about 4 months), which in turn is shorter than HSV vectors. The particular vector chosen will depend upon the target cell that one wants to look at a specific interaction in. The introduction can be by standard techniques, e.g. infection, transfection, transduction or transformation. Examples of modes of gene transfer include e.g., naked DNA, $CaPO_4$ precipitation, DEAE dextran, electroporation, protoplast fusion, lipofection, cell microinjection, and infection using viral vectors.

The vector can be employed to target essentially any desired target cell, such as a glioma. For example, stereotaxic injection can be used to direct the vectors (e.g. adenovirus, HSV) to a desired location. Additionally, the particles can be delivered by intracerebroventricular (icv) infusion using a minipump infusion system, such as a SynchroMed Infusion System. A method based on bulk flow, termed convection, has also proven effective at delivering large molecules to extended areas of the brain and may be useful in delivering the vector to the target cell (Bobo et al., *Proc. Natl. Acad. Sci. USA* 91:2076-2080

(1994); Morrison et al., *Am. J. Physiol.* 266:292-305 (1994)). Other methods that can be used include catheters, intravenous, parenteral, intraperitoneal and subcutaneous injection, and oral or other known routes of administration.

The cell to be tested, e.g. a knock-out cell (2 x 10^6 cells) can be transfected with DEAE-dextran using 10-14 μg of a vector such as pRc/CMV-sFv containing protein vector, and incubated at, for example, 37°C in RPMI media. 48 hours post-transfection, the cells are exposed to selection media with, for example, 500 μg/ml G418. Six to eight days later, the bulk stable cells are thoroughly washed with PBS. Cells are preferably grown at 37°C in a humidified incubator with 5% CO_2.

Preferably one transiently transfects the cells by known means such as described by Dean, et al., *Proc. Natl. Acad. Sci.* USA 90:8392-96 (1993).

Alternatively, the gene encoding the fusion protein containing the sFv can be cloned into a retroviral vector such as the LNCX MuLV shuttle vector under the control of the CMVIE promoter (A. Miller, Central Topics in Microbiology and Immunology, 158 (1991)). The vectors (10μg) can be transfected by calcium phosphate into a ecotropic cell line such as PE501 (10^6 cells/100 mm dishes) (A. Miller, Central Topics in Microbiology and Immunology, 158 (1991)). Twelve hours later, the cells are washed with PBS and three ml of fresh medium are added to the cells. After an additional 24 hours, the supernatants from the transfected cells are collected, cleared by low speed centrifugation (3000 x g; 1200 rpm), filtered through a 0.45 μm filter and three ml are used to infect the amphotropic packaging cell line, PG13 (10^6 cells/100 mm dish) in the presence of 8μg/ml protamine sulfate. 48 hours post-infection, the cells are washed and treated with selection medium containing 800μg/ml G418. Once producer cell lines are established, confluent monolayer cells are split and fresh medium is added. The cells are then incubated at 32°C, fusion protein containing supernatants are harvested, filtered and analyzed. One can transfect the cells with one of the fusion proteins and then transfect the cell lines with the other fusion protein or co-transfect the cells with both proteins. With certain vectors

such as herpes virus, HIV, pox virus, the cell can encode both fusion proteins in the same vector. One can also use transient transfection by known techniques. [Pear, W., et al., *Proc. Natl. Acad. Sci. USA*, 90:8392-8396 (1993). Preferably, one would transfect the cell in bulk.

Typically when one is using a library one first transfects the cells with the gene encoding the known protein. Thereafter, the bulk cells with the property of interest are selected or wells containing the transfected cells are transfected by different proteins from the library. When one is testing a specific protein pair, one preferably co-transfects the cells or uses a single vector expressing both proteins.

EXAMPLES

Human 293% cells were cotransfected with plasmids encoding GST or GST fused to the parental 9E10 ScFv or the second-generation derivative, 3DX, with or without plasmids encoding N-WASP tagged with 6 copies of the Myc tag (Myc_6NWASP) for a single HA tag (HA-NWASP). Fig 1 shows the amino acid and nucleotide sequence of 3DX. Amino acid residues that differ from the parental 9E10 sFx are in bold and underlined, with the corresponding parental 9E10 amino acids shown below in parenthesis. Two days post-transfection, cells were lysed in buffer containing 1% Triton X-100 and GST proteins were recovered by binding to glutathione-agarose beads. Proteins bound to beads (GSH pulldown) (Figure 1A) or whole cell lysates (Figure 2B) were immunoblotted with a polyclonal antibody recognizing N-WASP. The positions of various N-WASP species are indicated by arrows.

All references described herein are incorporated herein by reference.